

PhageAcr: Identification of anti-CRISPR proteins

Christian Neitzel, Fernanda Vieira, Hugo Oliveira, Óscar Dias

Center of Biological Engineering, University of Minho, 4710-057 Braga, Portugal

Context and Motivation

Bacteria and bacteriophages are engaged in a co-evolutionary arms race, driving the evolution of sophisticated defense mechanisms on both sides. Among these defenses, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system has emerged as a novel mechanism, enabling bacteria to boycott viral intrusion by precisely targeting and degrading viral DNA (Asmamaw & Zawdie, 2021; Hale et al., 2009; Horvath & Barrangou, 2010).

Similarly, in response to the CRISPR-Cas defense, bacteriophages have also developed countermeasure mechanisms, namely Anti-CRISPR proteins (Acrs). These compact molecules, encoded by a specific group of bacteriophages, act as an inhibitor, targeting the host (bacteria) defense mechanisms, specifically, CRISPR-Cas systems. Acrs ensure the successful invasion and replication of bacteriophages within their bacterial hosts through inhibition of CRISPR-Cas mechanisms (Bondy-Denomy et al., 2013).

A comprehensive overview of Anti-CRISPR systems is crucial for advancing genetic engineering and biotechnology, for example medical treatments (Tavakoli et al., 2021) and agriculture (Liu et al., 2021). This work aims to contribute to unveil insights about Anti-CRISPR systems, by fulfilling two main objectives. Firstly, it seeks to build a dataset of Acr proteins sequences from publicly available data such as NCBI. Secondly, it aims towards the exploration of machine learning approaches to enhance the accuracy of Acr predictions.

To obtain a better understanding of the mechanisms through which CRISPR-Cas defends bacteria in the inhibition of phage infection and how Acrs try to counter-act these mechanisms, we must study and develop bioinformatics tools based on: Machine Learning models; refinement of Acr nomenclature (Bondy-Denomy et al., 2018) and mining existing databases of known Acrs. Overall, these methods will allow not only to provide valuable insights of Acrs but also pave the way for further applications in the fields of microbial ecology and genome engineering.

1. Bacterial defense system

Bacteriophages (phages) are the most abundant viruses on the planet, infecting and replicating within bacteria by attaching to specific receptors on the bacterial surface and injecting their genetic material into the host cell. Following infection, phages may undergo either a lytic cycle, where they replicate within the host cell and ultimately cause cell lysis, releasing new phages, or a lysogenic cycle, in which the phage genome, known as the prophage, integrates into the bacterial genome. In the lysogenic cycle, the

integrated prophage can replicate along with the bacterial genome and remain latent until exposure to stress or environmental cues induces the prophage to enter the lytic cycle and produce new phage particles (Campbell, 2003; M. Zhang et al., 2022).

Phages play crucial roles in regulating bacterial populations and maintaining microbial diversity in several ecosystems. Additionally, phages gained attention for their potential applications in biotechnology, including phage therapy for combating bacterial infections and, as tools for genetic engineering (Guo et al., 2021; Haq et al., 2012).

To tackle the selective pressure of phages in bacterial communities, these last have urged to develop a wide range of defense mechanisms (Koskella & Brockhurst, 2014). Previous works point to this defense adaption as the so called prokaryotic immune system (Bernheim & Sorek, 2020). As part of this system we highlight: the restriction modification (RM) system, which cleaves and degrades viral DNA (Rodic et al., 2017); the abortive infection (Abi) system, which triggers host cell death to stop phage replication (Aframian & Eldar, 2023); and, the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR-Cas) system (Xu & Li, 2020).

In fact, the CRISPR-Cas system has gained significant attention, mainly in genetic engineering, due to its genome editing capabilities (Ran et al., 2013; Wei & Li, 2023). CRISPR-Cas or CRISPR are part of a family of proteins that act as adaptive immune systems, through the recognition and degradation of specific viral genetic information. This system is composed of two sections: a CRISPR array and Cas (CRISPR-associated) genes (Makarova et al., 2015). The CRISPR array consists of direct repeat sequences interspersed with variable-length "spacers." These spacers originate from fragments of prophage DNA that were cleaved and modified into "protospacers" by Cas proteins. Subsequently, these protospacers integrate into the CRISPR array, augmenting the host's defense mechanisms against similar phages that may attempt future infections (Sternberg et al., 2016). Cas genes encompass a cluster of genes responsible for encoding Cas proteins, which are essential for CRISPR-Cas defense mechanisms and information processing within the CRISPR array (Alkhnbashi et al., 2020; Hale et al., 2012).

CRISPR-Cas systems encompass three main stages: adaptation, expression and interference. The adaptation stage, Cas proteins identify a specific motif, known as a Protospacer-Adjacent Motif (PAM), on target DNA and remove a segment known as the protospacer for integration into the CRISPR array, serving as a spacer for future immunity. Some systems use an alternative mechanism, acquiring spacers from RNA via reverse transcription, thereby enhancing their adaptability. During the Expression stage, the CRISPR array is transcribed into a single pre-CRISPR RNA (pre-crRNA), which matures into CRISPR RNAs (crRNAs). These crRNAs, containing spacer sequences and flanking repeats, facilitate target recognition and interference. In the Interference stage, crRNA guides the CRISPR-Cas complex to recognize and cleave the protospacer or similar sequences within the invader's genome, leading to the inactivation of the invader's genetic material by Cas nucleases. This precise targeting provides a robust defense against viral infections and gene transfer events (Hampton et al., 2020).

2. CRISPR types and subtypes

CRISPR-Cas systems exhibit vast diversity in Cas protein sequences, gene compositions and genomic structures. This diversity suggests that these systems evolved and diversified across different species and environments, possibly as responses to the various pathogens encountered by the bacteria and archaea in their habitats. To further understand this diversity, researchers developed a classification of CRISPR-Cas systems based on their evolutionary relationships through comparisons of gene compositions, locus architectures, sequence similarity-based clustering and phylogenetic analysis of conserved Cas proteins (Makarova et al., 2015; Makarova & Koonin, 2015). This classification identified two primary classes, Class 1 and Class 2, each containing three distinct types with their own respective subtypes (Makarova et al., 2020). Class 1 systems feature effector modules comprising multiple Cas proteins that form a complex binding with crRNA, working together to bind and process the target. In contrast, class 2 systems are characterized by a single multidomain crRNA-binding protein, serving a functionally similar role to the entire effector complex found in class 1 systems (Makarova et al., 2020). Within these two classes, we can further classify distinct types based on the presence of unique signature Cas genes (Makarova & Koonin, 2015; S. Makarova et al., 2011), such as Cas3 being unique in type I, Cas9 in type II, Cas10 in type III, etc. The classification of subtypes, are assigned to systems that share most of the characteristics of a type but exhibit some minor deviations, warranting their classification as subtypes. In summary, recent publications define Class 1 as having 3 types (Types I, III, and IV) and 17 subtypes, while Class 2 consists of 3 types (Types II, V, VI) and 33 subtypes (S. Makarova et al., 2011).

2.1. Tools to detect CRISPR systems

Several tools were developed to assist in the identification and classification of CRISPR-Cas systems, namely CRISPRloci (Alkhnbashi et al., 2021), CRISPRFinder (Grissa et al., 2007) and CRISPRCasFinder (Couvin et al., 2018).

CRISPRloci uses machine learning algorithms to predict and assess CRISPR loci accurately, offering precise annotation of Cas genes and comprehensive insights into CRISPR array features like orientation, leader sequence, and Cas subtype classification, simplifying the characterization of CRISPR-Cas systems (Alkhnbashi et al., 2021).

CRISPRFinder is a web tool that specializes in identifying CRISPR-like structures in genomic sequences. characterized by its ability to define Direct Repeats (DRs), extracting spacers, and detecting smaller CRISPR candidates. Unlike earlier tools that often dismissed smaller CRISPR suspects (one or two motifs in size), CRISPRFinder's inclusive approach sheds light on CRISPR evolution and spread. While there's a risk of false candidates being included, CRISPRFinder addresses this by labeling them as "questionable," facilitating future refinement and accurate identification (Grissa et al., 2007).

CRISPRCasFinder, an upgraded version of CRISPRFinder, combines the detection of both CRISPR arrays and Cas proteins. Its enhanced CRISPR array detection tool shows superior performance in identifying Cas proteins, aligning with more recent classification

schemes (Couvin et al., 2018). Notably, CRISPRCasFinder not only boosts performance but also consolidates functions from other tools for comprehensive CRISPR identification.

3. Phage Anti-CRISPR

Bacteria and archaea constantly evolve anti-viral defenses, initiating an arms race with phages. Phages deploy Anti-CRISPR proteins (Acrs) to counteract CRISPR-Cas systems, inhibiting bacterial immune responses. Injected with viral DNA during infection, Acrs target Cas proteins, hindering DNA cleavage and facilitating infection (Bondy-Denomy et al., 2013). Acrs are short proteins, under 200 amino acids in length, often adopting compact, globular structures that allow them to effectively inhibit CRISPR-Cas activity (Pawluk, Staals, et al., 2016). Because of their ongoing co-evolution, bacteria and archaea have developed a more robust and diverse range of CRISPR-Cas systems, prompting phages to expand their Acr arsenal, with each subtype of CRISPR-Cas being targeted by specific Acrs (Koonin et al., 2017; Pawluk et al., 2018).

CRISPR-Cas has been harnessed for gene editing and synthetic gene circuit construction. Acrs, by acting as OFF-switches for CRISPR-Cas activity, offer benefits in the development of CRISPR-Cas based biotechnological tools, holding potential for improving gene editing techniques and overcoming challenges in eukaryotic synthetic gene circuits (Yu & Marchisio, 2020).

3.1. Databases for Anti-CRISPR

The continuous discovery of new Acr protein families makes it important to establish a standardized nomenclature to ensure clarity and consistency among researchers. Efforts towards this goal, as referenced in (Bondy-Denomy et al., 2015; Pawluk, Amrani, et al., 2016), have laid the groundwork for a unified Acr naming approach. Building upon this foundation, researchers have established a Google Doc database (accessible via <https://tinyurl.com/anti-CRISPR>), serving as a comprehensive repository for cataloging newly identified Acr proteins and ensuring conformity to established naming conventions (Bondy-Denomy et al., 2018).

To accelerate and broaden the exploration of Acrs, user-friendly and regularly updated databases have emerged such as Anti-CRISPRdb (Dong et al., 2018, 2022) and CRISPRminer (F. Zhang et al., 2018). Leveraging Python scripts, Anti-CRISPRdb (available at <http://guolab.whu.edu.cn/anti-CRISPRdb/>) extracts data from PubMed and Google Scholar, followed by sequence alignment through NCBI's BLAST. Its primary aim is to streamline researchers' efforts by providing comprehensive functions for searching, browsing, downloading and uploading data on anti-CRISPR proteins. With accessible information encompassing family, function, source species, verification status, comments and references, researchers can efficiently navigate through the database. The intuitive interface, coupled with robust search capabilities and BLAST integration, facilitates swift access to Acr proteins and their potential candidates, thus fostering accelerated research progress in the field (Dong et al., 2018). CRISPRminer (available at <http://www.microbiome-bigdata.com/CRISPRminer>) on the other hand provides

advanced functionalities for CRISPR-Cas system analysis, provides (as of late) information on 23 families of Acrs, and offers extensive analysis tools. In short, Anti-CRISPRdb and CRISPRminer have unique advantages, and they can complement each other to provide insights into Acr proteins and CRISPR-Cas systems.

3.2. Tools to detect Anti-CRISPR

Identifying Acrs poses a significant challenge due to the variability in their amino acid sequences (Pawluk et al., 2018), reflecting the diverse nature of CRISPR-Cas systems. Their compact size and evolutionary diversity complicate detection, hindering traditional sequence analysis methods. While conventional tools relied on guilt-by-association (GBA) and self-targeting methods, Machine Learning approaches can accelerate Acr identification. Training the model on databases of confirmed Acrs and non-Acrs, it learns distinguishing characteristics and discerns features that differentiate Acrs from non-Acrs, enhancing the identification of candidate Acr families (A. B. Gussow et al., 2020).

AcrFinder (available at <http://bcb.unl.edu/AcrFinder>), integrates a multifaceted and well-accepted approach, combining homology search, GBA and CRISPR-Cas self-targeting spacers, allowing for increased likelihood of identifying Acrs (Yi et al., 2020).

Homology search identifies Acr homologs through sequence similarity, while GBA identifies Aca proteins co-localized with Acrs, indicating their functional association. Additionally, analysis of CRISPR-Cas self-targeting spacers validates the presence of Acr-Aca operons, highlighting interactions between Acr proteins and the CRISPR-Cas systems.

AcRanker (<https://bio.tools/AcRanker>) introduces a novel method for identifying potential Acrs directly from protein sequences (Eitzinger et al., 2020), using the XGBoost algorithm (Chen & Guestrin, 2016). XGBoost is a powerful machine learning algorithm known for its exceptional predictive accuracy. It utilizes gradient boosting to optimize the loss function's gradient and integrates regularization techniques to prevent overfitting. Through an ensemble of weak decision trees built sequentially, XGBoost corrects errors made by previous trees, refining predictions iteratively until reaching high accuracy. The development of AcRanker used XGBoost as a pairwise ranking model to rank proteins within a proteome based on their expected Acr behavior by using only 1-, 2- and 3-mer amino acid composition as input features. The model was trained with a dataset of 20 experimentally verified Acrs taken from the anti-CRISPRdb database, subsequently testing was done of an additional set of 20 known Acrs. From their study, it was discovered that ML rankings for Acrs contained in phage proteomes were much better than those in bacterial proteomes, possibly due to their smaller size. AcRanker streamlines the identification process, enabling efficient ranking of potential Acr candidate genes for further testing and validation. Notably, the tool discovered two previously unknown Acrs, AcrIIA20 (ML1) and AcrIIA21 (ML8), demonstrating its efficiency in Acr detection. Additionally, AcRanker offers a complementary approach for detecting Acr proteins, providing a valuable tool for researchers in the field.

AcrCatalog (<https://acrcatalog.pythonanywhere.com/catalog/>) is a study and resource dedicated to identifying Acr proteins, employing machine learning and bioinformatic techniques to predict and characterize novel Acrs (A. Gussow et al., 2020). The predictive

model utilized by AcrCatalog is based on the random forest algorithm, a machine learning approach that consists of an ensemble of decision trees. In this algorithm, each decision tree is trained with random parts of the training data to reach a single result (Breiman, 2001). During the development of AcrCatalog, a random forest model with 1000 decision trees was employed, which identified 2500 previously undetected candidates. The methodology of AcrCatalog involves enriching search spaces within prokaryotic and viral genomes, clustering and weighting candidate Acrs, and applying heuristic filters.

Objectives

The main objective of this work is to develop a tool for identifying Acrs. Our aim is to build a comprehensive dataset of Acr proteins sequences from public available data such as NCBI. Furthermore, we plan to explore Machine Learning approaches to enhance the accuracy of Acr predictions, leveraging recent Acr databases and analyzing their methodologies. Our goal is to surpass the current state-of-the-art by developing a cutting-edge Machine Learning model for predicting Acrs. Lastly, we will conduct benchmarking of developed tools to evaluate and compare their performance and accuracy in predicting Acrs.

Bibliography

- Aframian, N., & Eldar, A. (2023). Abortive infection antiphage defense systems: Separating mechanism and phenotype. *Trends in Microbiology*, 31(10), 1003–1012. <https://doi.org/10.1016/j.tim.2023.05.002>
- Alkhnbashi, O. S., Meier, T., Mitrofanov, A., Backofen, R., & Voß, B. (2020). CRISPR-Cas bioinformatics. *Methods*, 172, 3–11. <https://doi.org/10.1016/j.ymeth.2019.07.013>
- Alkhnbashi, O. S., Mitrofanov, A., Bonidia, R., Raden, M., Tran, V. D., Eggenhofer, F., Shah, S. A., Öztürk, E., Padilha, V. A., Sanches, D. S., de Carvalho, A. C. P. L. F., & Backofen, R. (2021). CRISPRloci: Comprehensive and accurate annotation of CRISPR–Cas systems. *Nucleic Acids Research*, 49(W1), W125–W130. <https://doi.org/10.1093/nar/gkab456>
- Asmamaw, M., & Zawdie, B. (2021). Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. *Biologics : Targets & Therapy*, 15, 353–361. <https://doi.org/10.2147/BTT.S326422>
- Bernheim, A., & Sorek, R. (2020). The pan-immune system of bacteria: Antiviral defence as a community resource. *Nature Reviews Microbiology*, 18(2), 113–119. <https://doi.org/10.1038/s41579-019-0278-2>
- Bondy-Denomy, J., Davidson, A. R., Doudna, J. A., Fineran, P. C., Maxwell, K. L., Moineau, S., Peng, X., Sontheimer, E. J., & Wiedenheft, B. (2018). A Unified Resource for Tracking Anti-CRISPR Names. *The CRISPR Journal*, 1(5), 304–305. <https://doi.org/10.1089/crispr.2018.0043>
- Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K. L., & Davidson, A. R. (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature*, 526(7571), 136–139. <https://doi.org/10.1038/nature15254>

- Bondy-Denomy, J., Pawluk, A., Maxwell, K. L., & Davidson, A. R. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493(7432), 429–432. <https://doi.org/10.1038/nature11723>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Campbell, A. (2003). The future of bacteriophage biology. *Nature Reviews. Genetics*, 4(6), 471–477. <https://doi.org/10.1038/nrg1089>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E. P. C., Vergnaud, G., Gautheret, D., & Pourcel, C. (2018). CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research*, 46(Web Server issue), W246–W251. <https://doi.org/10.1093/nar/gky425>
- Dong, C., Hao, G.-F., Hua, H.-L., Liu, S., Labena, A. A., Chai, G., Huang, J., Rao, N., & Guo, F.-B. (2018). Anti-CRISPRdb: A comprehensive online resource for anti-CRISPR proteins. *Nucleic Acids Research*, 46(Database issue), D393–D398. <https://doi.org/10.1093/nar/gkx835>
- Dong, C., Wang, X., Ma, C., Zeng, Z., Pu, D.-K., Liu, S., Wu, C.-S., Chen, S., Deng, Z., & Guo, F.-B. (2022). Anti-CRISPRdb v2.2: An online repository of anti-CRISPR proteins including information on inhibitory mechanisms, activities and neighbors of curated anti-CRISPR proteins. *Database*, 2022, baac010. <https://doi.org/10.1093/database/baac010>
- Eitzinger, S., Asif, A., Watters, K. E., Iavarone, A. T., Knott, G. J., Doudna, J. A., & Minhas, F. ul A. A. (2020). Machine learning predicts new anti-CRISPR proteins. *Nucleic Acids Research*, 48(9), 4698–4708. <https://doi.org/10.1093/nar/gkaa219>
- Grissa, I., Vergnaud, G., & Pourcel, C. (2007). CRISPRFinder: A web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 35(Web Server issue), W52–W57. <https://doi.org/10.1093/nar/gkm360>
- Guo, D., Chen, J., Zhao, X., Luo, Y., Jin, M., Fan, F., Park, C., Yang, X., Sun, C., Yan, J., Chen, W., & Liu, Z. (2021). Genetic and Chemical Engineering of Phages for Controlling Multidrug-Resistant Bacteria. *Antibiotics*, 10(2), 202. <https://doi.org/10.3390/antibiotics10020202>
- Gussow, A. B., Park, A. E., Borges, A. L., Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Bondy-Denomy, J., & Koonin, E. V. (2020). Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nature Communications*, 11(1), 3784. <https://doi.org/10.1038/s41467-020-17652-0>
- Gussow, A., Shmakov, S., Makarova, K., Wolf, Y., Bondy-Denomy, J., & Koonin, E. (2020). *Vast diversity of anti-CRISPR proteins predicted with a machine-learning approach*. <https://doi.org/10.1101/2020.01.23.916767>
- Hale, C. R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A. M., Glover, C. V. C., Graveley, B. R., Terns, R. M., & Terns, M. P. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Molecular Cell*, 45(3), 292–302. <https://doi.org/10.1016/j.molcel.2011.10.023>
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., & Terns, M. P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, 139(5), 945–956. <https://doi.org/10.1016/j.cell.2009.07.040>
- Hampton, H. G., Watson, B. N. J., & Fineran, P. C. (2020). The arms race between bacteria and their phage foes. *Nature*, 577(7790), 327–336. <https://doi.org/10.1038/s41586-019-1894-8>
- Haq, I. U., Chaudhry, W. N., Akhtar, M. N., Andleeb, S., & Qadri, I. (2012). Bacteriophages and their implications on future biotechnology: A review. *Virology Journal*, 9, 9. <https://doi.org/10.1186/1743-422X-9-9>

- Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science (New York, N.Y.)*, 327(5962), 167–170. <https://doi.org/10.1126/science.1179555>
- Koonin, E. V., Makarova, K. S., & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*, 37, 67–78. <https://doi.org/10.1016/j.mib.2017.05.008>
- Koskella, B., & Brockhurst, M. A. (2014). Bacteria–phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *Fems Microbiology Reviews*, 38(5), 916–931. <https://doi.org/10.1111/1574-6976.12072>
- Liu, Q., Yang, F., Zhang, J., Liu, H., Rahman, S., Islam, S., Ma, W., & She, M. (2021). Application of CRISPR/Cas9 in Crop Quality Improvement. *International Journal of Molecular Sciences*, 22(8), 4206. <https://doi.org/10.3390/ijms22084206>
- Makarova, K. S., & Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems. *Methods in Molecular Biology (Clifton, N.J.)*, 1311, 47–75. https://doi.org/10.1007/978-1-4939-2687-9_4
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Terns, R. M., Terns, M. P., White, M. F., Yakunin, A. F., Garrett, R. A., van der Oost, J., ... Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, 13(11), 722–736. <https://doi.org/10.1038/nrmicro3569>
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., Terns, M. P., Venclovas, Č., White, M. F., Yakunin, A. F., ... Koonin, E. V. (2020). Evolutionary classification of CRISPR-Cas systems: A burst of class 2 and derived variants. *Nature Reviews. Microbiology*, 18(2), 67–83. <https://doi.org/10.1038/s41579-019-0299-x>
- Pawluk, A., Amrani, N., Zhang, Y., Garcia, B., Hidalgo-Reyes, Y., Lee, J., Edraki, A., Shah, M., Sontheimer, E. J., Maxwell, K. L., & Davidson, A. R. (2016). Naturally Occurring Off-Switches for CRISPR-Cas9. *Cell*, 167(7), 1829–1838.e9. <https://doi.org/10.1016/j.cell.2016.11.017>
- Pawluk, A., Davidson, A. R., & Maxwell, K. L. (2018). Anti-CRISPR: Discovery, mechanism and function. *Nature Reviews Microbiology*, 16(1), 12–17. <https://doi.org/10.1038/nrmicro.2017.120>
- Pawluk, A., Staals, R. H. J., Taylor, C., Watson, B. N. J., Saha, S., Fineran, P. C., Maxwell, K. L., & Davidson, A. R. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature Microbiology*, 1(8), 1–6. <https://doi.org/10.1038/nmicrobiol.2016.85>
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., & Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*, 8(11), 2281–2308. <https://doi.org/10.1038/nprot.2013.143>
- Rodic, A., Blagojevic, B., Zdobnov, E., Djordjevic, M., & Djordjevic, M. (2017). Understanding key features of bacterial restriction-modification systems through quantitative modeling. *BMC Systems Biology*, 11(Suppl 1), 1–15. <https://doi.org/10.1186/s12918-016-0377-x>
- S. Makarova, K., H. Haft, D., Barrangou, R., J. J. Brouns, S., Charpentier, E., Horvath, P., Moineau, S., J. M. Mojica, F., I. Wolf, Y., Yakunin, A. F., van der Oost, J., & V. Koonin, E. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews. Microbiology*, 9(6), 467–477. <https://doi.org/10.1038/nrmicro2577>
- Sternberg, S. H., Richter, H., Charpentier, E., & Qimron, U. (2016). Adaptation in CRISPR-Cas Systems. *Molecular Cell*, 61(6), 797–808. <https://doi.org/10.1016/j.molcel.2016.01.030>
- Tavakoli, K., Pour-Aboughadareh, A., Kianersi, F., Poczai, P., Etminan, A., & Shooshtari, L. (2021). Applications of CRISPR-Cas9 as an Advanced Genome Editing System in Life Sciences. *BioTech*, 10(3), 14. <https://doi.org/10.3390/biotech10030014>

- Wei, J., & Li, Y. (2023). CRISPR-based gene editing technology and its application in microbial engineering. *Engineering Microbiology*, 3(4), 100101. <https://doi.org/10.1016/j.engmic.2023.100101>
- Xu, Y., & Li, Z. (2020). CRISPR-Cas systems: Overview, innovations and applications in human disease research and gene therapy. *Computational and Structural Biotechnology Journal*, 18, 2401–2415. <https://doi.org/10.1016/j.csbj.2020.08.031>
- Yi, H., Huang, L., Yang, B., Gomez, J., Zhang, H., & Yin, Y. (2020). AcrFinder: Genome mining anti-CRISPR operons in prokaryotes and their viruses. *Nucleic Acids Research*, 48(W1), W358–W365. <https://doi.org/10.1093/nar/gkaa351>
- Yu, L., & Marchisio, M. A. (2020). Types I and V Anti-CRISPR Proteins: From Phage Defense to Eukaryotic Synthetic Gene Circuits. *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.575393>
- Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., Zhou, F., Jia, Y., Zheng, K., & Huang, Z. (2018). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Communications Biology*, 1, 180. <https://doi.org/10.1038/s42003-018-0184-6>
- Zhang, M., Zhang, T., Yu, M., Chen, Y.-L., & Jin, M. (2022). The Life Cycle Transitions of Temperate Phages: Regulating Factors and Potential Ecological Implications. *Viruses*, 14(9), 1904. <https://doi.org/10.3390/v14091904>